

**COGNITION
ET TRAITEMENT
DE L'INFORMATION**

Information - Commande - Communication

3

Sémantique et traitement automatique du langage naturel

*sous la direction de
Patrice Enjalbert*

hermes

Lavoisier

Table des matières

Introduction	17
PREMIÈRE PARTIE. REPÈRES	25
Chapitre 1. Sémantique et traitement automatique du langage naturel : première approche	27
Patrice ENJALBERT	
1.1. Qu'est-ce que « comprendre un texte » ?	27
1.2. La construction du sens	32
1.2.1. Construction de « représentations »	32
1.2.2. Par quels procédés, avec quelles connaissances ?	34
1.2.3. Et pour une « machine » ? Compréhension automatique limitée de corpus homogènes	38
1.3. Exemples d'applications « à composante sémantique »	40
1.3.1. Une tâche d'école : compréhension automatique et visualisation de constats d'accidents	40
1.3.2. Extraction d'information	42
1.3.3. Recherche d'information « par le contenu »	43
1.3.4. Dialogue oral : le système Artimis	45
1.3.5. Interaction multimodale pour la recherche d'informations géographiques	47
1.4. Conclusion	49
1.5. Bibliographie	50
Chapitre 2. Les paliers de la sémantique	53
Patrice ENJALBERT et Bernard VICTORRI	
2.1. Le mot	54
2.1.1. Quelles unités minimales ?	54
2.1.2. Représentations du sens lexical	57

2.1.2.1. Des relations entre concepts	57
2.1.2.2. Entités, propriétés et relations	60
2.1.2.3. Sémantique différentielle (ou componentielle)	61
2.1.3. Calcul du sens lexical	64
2.1.4. Conclusion : quels modèles pour le TALN ?	67
2.2. La phrase	68
2.2.1. Caractérisation du niveau : intégration sémantique	68
2.2.2. Instanciation	70
2.2.3. Relations argumentales et compositionnalité.	72
2.2.4. Repérage, cadrage et opérations énonciatives	74
2.3. Le texte	76
2.3.1. Structure et cohésion textuelle	76
2.3.2. Coréférence	77
2.3.3. Temporalité	79
2.3.4. Structures du récit.	80
2.3.5. Isotopies	81
2.3.6. Structuration thématique.	82
2.4. La pragmatique.	84
2.4.1. La pragmatique comme étude de l'usage de la langue	84
2.4.2. Énonciation et actes de langage.	85
2.4.3. Interprétation (1) : sens littéral et implicature	86
2.4.4. Interprétation (2) : « le lecteur modèle »	89
2.4.5. Conclusion. Un « au-delà » de la sémantique ?	91
2.5. Bibliographie.	93

DEUXIÈME PARTIE. MODÉLISATION SÉMANTIQUE. 97

Chapitre 3. Polysémie lexicale 99

Guillaume JACQUET, Fabienne VENANT, Bernard VICTORRI

3.1. Introduction.	99
3.2. Les différentes formes de polysémie	101
3.2.1. Polysémie et homonymie : continuités et ruptures de sens.	101
3.2.2. Phénomènes de parole et degrés de lexicalisation	103
3.2.3. Métonymies intégrées et coercition de type	105
3.2.4. Représentations de la polysémie lexicale	108
3.3. Une représentation géométrique de la polysémie.	111
3.3.1. Représentation des différents sens d'une unité polysémique	112
3.3.2. Construction de l'espace sémantique à l'aide d'un graphe de synonymie	114
3.4. Calcul du sens d'une unité polysémique dans un énoncé donné.	118
3.4.1. Principes de la méthode de calcul du sens	119

3.4.2. Constructions syntaxiques et classes de sélection distributionnelle	122
3.4.3. Conclusion	128
3.5. Bibliographie	129
Chapitre 4. Le calcul de la référence	133
Bernard VICTORRI	
4.1. Présentation	133
4.1.1. Notion de chaîne de coréférence	133
4.1.2. Cadre de cette étude	136
4.2. Les différentes approches	139
4.2.1. Heuristiques, contraintes syntaxiques et sémantiques	139
4.2.2. Processus inférentiels et représentation de la connaissance	140
4.2.3. La notion de focus	141
4.2.4. La théorie du centrage	144
4.2.5. Modèle du contexte et saillance	146
4.3. Un nouveau modèle de l'identification des entités	148
4.3.1. Le modèle des attentes	149
4.3.2. Le calcul de la saillance	152
4.3.3. Plage de saillance et concordance	154
4.3.4. L'identification des entités	158
4.4. Un système d'expérimentation : Calcoref	161
4.4.1. Un logiciel voué à l'expérimentation	161
4.4.2. L'architecture générale	164
4.4.3. Premiers résultats	166
4.5. Conclusion	168
4.6. Bibliographie	170
Chapitre 5. Temporalité	173
Laurent GOSSELIN et Cédric PERSON	
5.1. Introduction	173
5.2. Présentation du modèle	178
5.2.1. Le format de représentation	178
5.2.1.1. Le point de référence	178
5.2.1.2. Des points aux intervalles	180
5.2.1.3. Définitions	183
5.2.2. Les principes de calcul : la compositionnalité holiste	188
5.2.2.1. Instructions et conflits	188
5.2.2.2. Anaphore et prise en compte de la dimension textuelle	191
5.2.3. Conclusion : extensions linguistiques du modèle	197
5.3. Implémentation : le système TNLP	198

5.3.1. Introduction	198
5.3.2. Marqueurs aspectuo-temporels et représentation intermédiaire du texte	201
5.3.2.1. Marques internes à la proposition	202
5.3.2.2. Dimension textuelle	202
5.3.2.3. Relations pragmatico-référentielles	203
5.3.3. Application des règles	203
5.3.3.1. Principe d'application	203
5.3.3.2. Règles compositionnelles et résolutions de conflits	204
5.3.3.3. Dimension textuelle et cohérence temporelle	205
5.3.4. Exemple de traitement	206
5.3.5. Conclusion	209
5.4. Bibliographie	211

Chapitre 6. Sémantique de l'espace et du déplacement 215

Yann MATHET

6.1. Circonscrire le domaine spatial	216
6.1.1. Un lexique spatial ?	216
6.1.2. Un sens spatial ?	218
6.1.2.1. Autonomie du domaine spatial ?	218
6.1.2.2. Un entrelacement privilégié : l'espace et le temps	220
6.1.3. Bilan	220
6.2. Cahier des charges d'une étude sur la spatio-temporalité en langue.	221
6.2.1. Introduction	221
6.2.2. Observations préliminaires sur la spatio-temporalité telle qu'elle se manifeste en langue.	222
6.2.3. Une première réponse à la recherche de référents spatiaux : la localisation par des lieux	222
6.2.4. Mise au jour d'autres paradigmes que les lieux	225
6.2.4.1. La distance	225
6.2.4.2. L'orientation et la direction	225
6.2.4.3. Notion de « forme »	226
6.2.5. Autonomie de la notion de déplacement	228
6.2.6. Clivage statique <i>versus</i> déplacement ?	229
6.2.6.1. Relations de forme	230
6.2.6.2. Mixité statique/déplacement	231
6.2.6.3. Les états vont-ils de pair avec la description du statique, et réciproquement ?	231
6.2.7. Illustration sur quelques exemples des types de relations mises en jeu par des expressions spatio-temporelles	232
6.2.8. Conclusion	234
6.3. Proposition : un modèle à plusieurs paradigmes	236

6.3.1. Les entités	236
6.3.2. Polymorphisme sur les entités	238
6.3.3. Les relations	238
6.3.3.1. Relations topologiques	239
6.3.3.2. Relations de distance.	241
6.3.3.3. Relations de forme	242
6.3.3.4. Relations d'ordre curviligne	244
6.4. Etude du lexique spatio-temporel s'appuyant sur le modèle	244
6.4.1. Etude « locale » du lexique : adéquation d'une acception (verbe, préposition) avec une configuration spatio-temporelle donnée.	244
6.4.1.1. Implémentation des entités du modèle	245
6.4.1.2. Implémentation du polymorphisme sur les entités	246
6.4.1.3. Implémentation des relations	247
6.4.1.4. Exploitation, résultats	249
6.4.2. Etude « macroscopique » du lexique : vers une approche componentielle	250
6.4.2.1. Contexte et objectifs	250
6.4.2.2. Constitution d'un graphe du domaine.	251
6.4.2.3. Observation du verbe <i>contourner</i>	253
6.4.2.4. Comparaison de l'expansion sur le graphe de deux verbes : <i>entrer</i> et <i>aller</i>	254
6.4.2.5. Utilisation du graphe, perspectives de recherche	255
6.4.3. Collecte de données psycholinguistiques pour l'étude et l'évaluation en TALN	256
6.5. Bibliographie	257

TROISIÈME PARTIE. DE LA COMPRÉHENSION AUTOMATIQUE AUX APPLICATIONS DOCUMENTAIRES

261

Chapitre 7. Compréhension automatique

267

Thierry CHARNOIS et Patrice ENJALBERT

7.1. Points de repère historiques et principes généraux des systèmes de compréhension automatique	268
7.1.1. Repères historiques	268
7.1.2. Principes des systèmes de compréhension automatique	271
7.2. Représentation et calcul du sens de la phrase en logique	273
7.2.1. La logique du premier ordre.	273
7.2.2. Représentation et calcul du sens d'un énoncé	276
7.2.2.1. Principe général et λ -notation	276
7.2.2.2. Représentation du lexique et des éléments grammaticaux	279
7.2.2.3. Sémantique compositionnelle de la phrase.	282
7.3. Représentation et calcul du sens de la phrase en graphes conceptuels	284

Chapitre 9. L'accès assisté à l'information documentaire	335
Patrice ENJALBERT et Frédéric BILHAUT	
9.1. De l'indexation manuelle à l'annotation sémantique automatique : aperçu général	336
9.2. Les moteurs de « première génération »	338
9.2.1. Nature des descripteurs	339
9.2.2. Méthode d'indexation et de recherche	340
9.2.3. Moteurs de « première génération et demi »	342
9.2.4. Discussion	344
9.3. Recherche d'information par requêtes structurées : entre recherche documentaire et extraction d'information	346
9.3.1. Recherche documentaire structurée	347
9.3.2. Systèmes de question/réponse (<i>Question Answering</i>)	348
9.3.3. Discussion	351
9.4. A la recherche de la structure du document : du résumé à la « navigation intradocumentaire »	353
9.4.1. Segmentation et structuration thématique	354
9.4.2. Résumé automatique	356
9.4.3. Liens hypertextuels	358
9.4.4. Discussion	359
9.5. Le Web sémantique	360
9.6. Conclusion	363
9.7. Bibliographie	364
 Chapitre 10. Recherche d'information géographique	 371
Frédéric BILHAUT et Patrice ENJALBERT	
10.1. Objectifs et architecture d'ensemble d'un système de recherche d'information géographique	372
10.1.1. Corpus et tâche	372
10.1.2. Architecture	374
10.2. Extraction des descripteurs à partir des expressions spatiales et temporelles	376
10.2.1. Expressions spatiales	376
10.2.1.1. Structure syntactico-sémantique	376
10.2.1.2. Représentations sémantiques	378
10.2.1.3. Implémentation	379
10.2.2. Expressions temporelles	380
10.2.2.1. Description des expressions et principe de l'analyse	380
10.2.2.2. Exemple	381
10.2.2.3. Mise en œuvre, discussion	383
10.3. Analyse discursive	383

10.3.1. Sélection de passages textuels : mise en relation des trois types d'expressions.	383
10.3.2. Cadres de discours spatiaux et temporels	384
10.3.2.1. La théorie de l'encadrement du discours	384
10.3.2.2. Détection et délimitation des cadres de discours.	385
10.3.2.3. Conclusion	387
10.4. Analyse thématique multidimensionnelle	388
10.4.1. Introduction	388
10.4.2. Thèmes composites et axes sémantiques	389
10.4.2.1. La notion de thème composite	389
10.4.2.2. Autres configurations discursives	391
10.4.3. Méthode d'analyse automatique	393
10.4.3.1. Analyse automatique en thèmes composites	393
10.4.3.2. Extraction automatique d'axes sémantiques	394
10.4.4. Conclusion	396
10.5. Conclusion	397
10.6. Annexe : la plate-forme <i>LinguaStream</i>	400
10.7. Bibliographie	403
Index	407